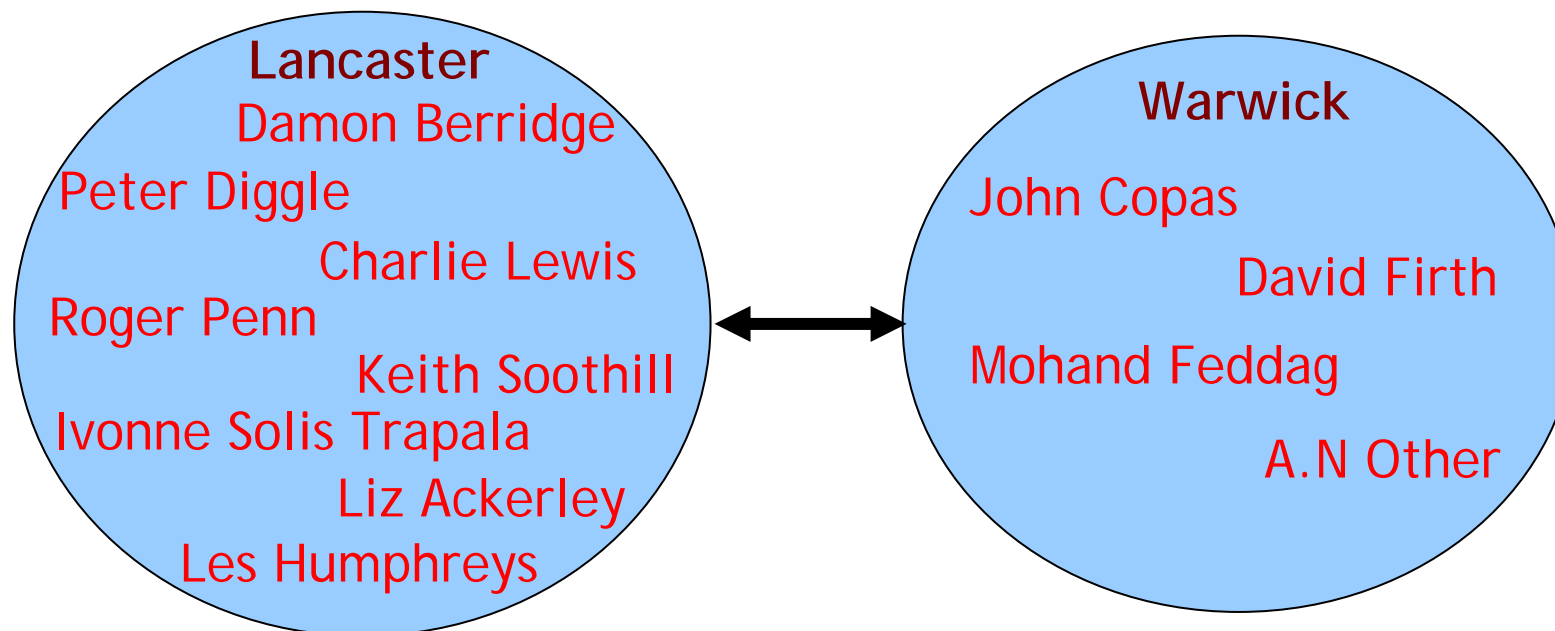


The Lancaster-Warwick node

Developing Statistical Modelling in the Social Sciences

Director: Brian Francis (Lancaster)



... Introduction

Our broad research theme is the statistical modelling of social development and change.

The research is focussed on three major substantive projects which are specifically related to longitudinal data:

- ✱ **Sociology/Political science:** Value changes in society (Penn, Firth, Francis, Humphreys)
- ✱ **Criminology:** Patterns of criminal behaviour (Soothill, Francis, Copas, Ackerley, Humphreys)
- ✱ **Psychology:** Child development and microgenetic studies (Lewis, Diggle, Copas, Solis Trapala)

While each theme has its own particular statistical methodological issues, the themes are also related by statistical computational issues:

- ✱ **Statistical methodology:** the development of efficient computational methods for dealing with complex random effects (Firth, Feddag)

... Value changes in society

... Value changes in society

Modelling longitudinal ranked data and longitudinal likert scales - social values.

What is the issue?

Example: The Economist: Special report on American Values, Jan 2003.

“We share common values- the common values of freedom, human rights and democracy” George Bush, November, 2002 in the Czech Republic

“It is time to stop pretending that Europeans and Americans share a common view of the world - Americans are from Mars and Europeans are from Venus” Robert Kagan, Carnegie Endowment



... Value changes in society -debates

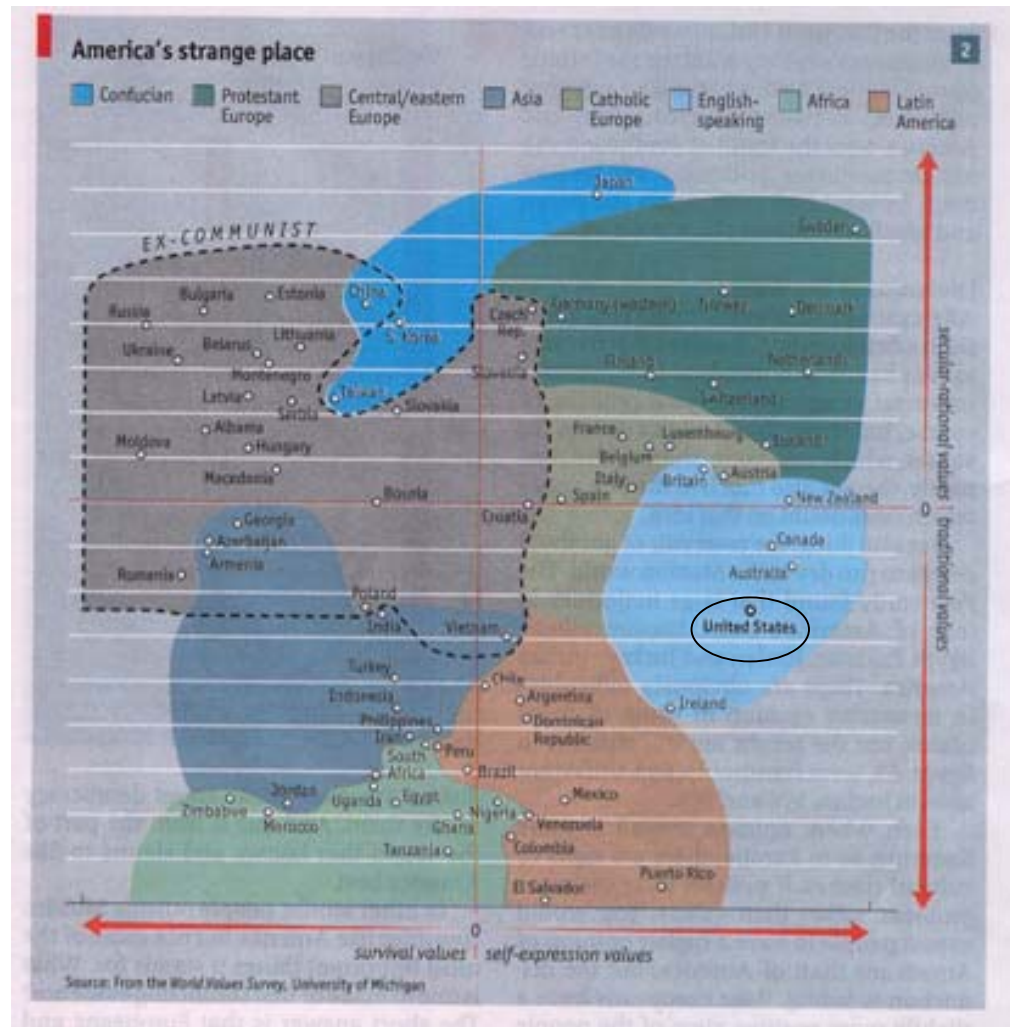
General homogenization of values:

- ✱ Convergence thesis: 1950s/1960s. C Kerr *et al* (1954) argued that values were 'converging' on the pattern of liberal democracy. 'End of Ideology' thesis embodied similar notion.
- ✱ Globalization theory: once again resurgence of views that there is a unilinear path to value orientations. F Fukuyama(1992)

The Persistence of Fundamental Differences in Societal Value systems:

- ✱ S. Huntington's 'Clash of Civilizations' (1993) argued for irrevocable differences in values. In particular he emphasized the clash between 'Western' values and those of Islam.
- ✱ Others (including Inglehart) have delineated a growing clash between 'secularism' and values associated with the Bible (Judaism/Christianity/Islam)

... Value changes in society - The Economist article



*world values
survey*

secular-
rational



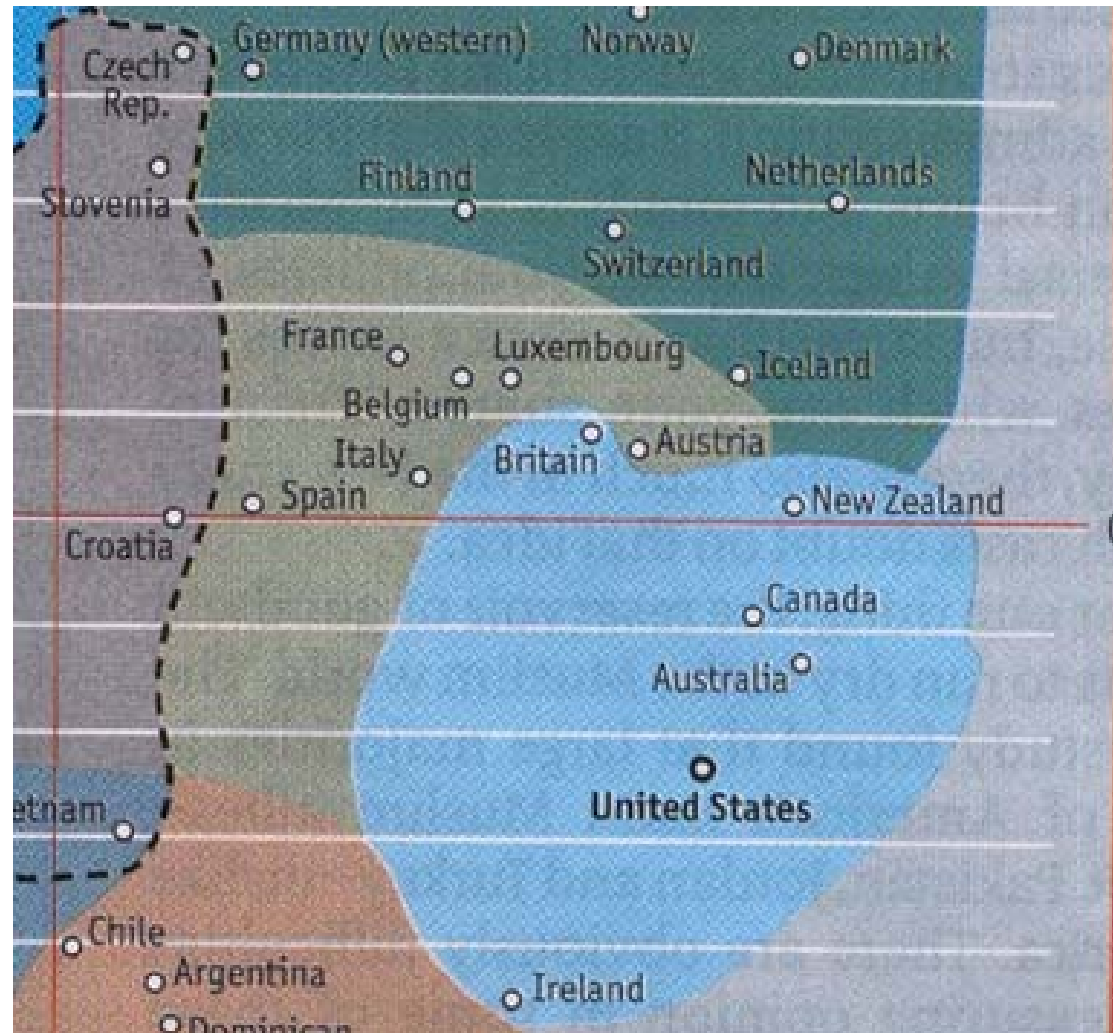
traditional

survival



self expression

... Value changes in society



... Value changes in society

Horizontal axis measures concept of materialism and postmaterialism which was proposed by Inglehart.

Background : Brecht : *“Zuerst das Fressen, dann die Moral”*
Maslow’s “hierarchy of needs”

Postmaterialist values are defined as:
“social equality, life style choices, participation, concern for environmental quality”

... Value changes in society

Question V7

Looking at the list below, please tick a box next to the one thing you think should be Britain's **highest priority**, the most important thing it should do.

Britain should ...

Highest
Priority

Items:

Maintain order in nation

1

Give people more to say in government decisions

2

Fight rising prices

3

Protect freedom of speech

4

Can't choose

8

... Value changes in society

Question V8

And which one do you think should be Britain's **next highest priority**, the second most important thing it should do?

Britain should ...

Next highest

Priority

Items:

Maintain order in nation	1
Give people more to say in government decisions	2
Fight rising prices	3
Protect freedom of speech	4
Can't choose	8

... Value changes in society

Questions come from Inglehart(1990) '*Cultural shift in Advanced Industrial Society*' to assess postmaterialist and materialist values.

Inglehart states that:

Those choosing items 2 and 4 as top two priorities are [post-materialist](#)
'more say in government' *'protect freedom of speech'*

Those choosing items 1 and 3 as top two priorities are [materialist](#)
'maintain order in the nation' *'Fight rising prices'*

All others are '[mixed](#)'.

1993 International Social Science Programmes survey.

Look at 5 countries - W and E Germany, Italy, UK and Poland.

Responses to the Inglehart Index questions for the five countries combined.

	Second priority						
First (top) priority	Maintain order	People more to say	Fight rising prices	Freedom of speech	Can't choose	Refused	Total
Maintain order	72	722	1110	347	42	4	2297
People more to say	699	73	594	407	26	4	1803
Fight rising prices	645	375	72	99	43	3	1237
Freedom of speech	149	168	70	4	5	4	400
Can't choose	15	8	14	4	93	0	134
Refused	7	1	5	2	0	30	45
Total	1587	1347	1865	863	209	45	5916



Materialists



Postmaterialists

... Value changes in society - some questions

- ✱ “Younger Europeans have been raised in relative prosperity, and are shifting towards postmaterial values”
- ✱ “Postmaterial values are less developed in Britain” Inglehart (1990,1997)
- ✱ Is this simplistic? What other covariates apart from country affect response on Inglehart’s scale? How does response vary with age? Is there a generational effect?

We wish to critically appraise the existing methodology of scale construction, and attempt to identify value shifts across nations. In particular, to examine longitudinal ranked and likert scale data, and to develop methods for analyzing such data.

Need to incorporate temporal dependence and non-linear changes for age and generational effects to identify whether such shifts are generational.

... Value changes in society

... some initial work

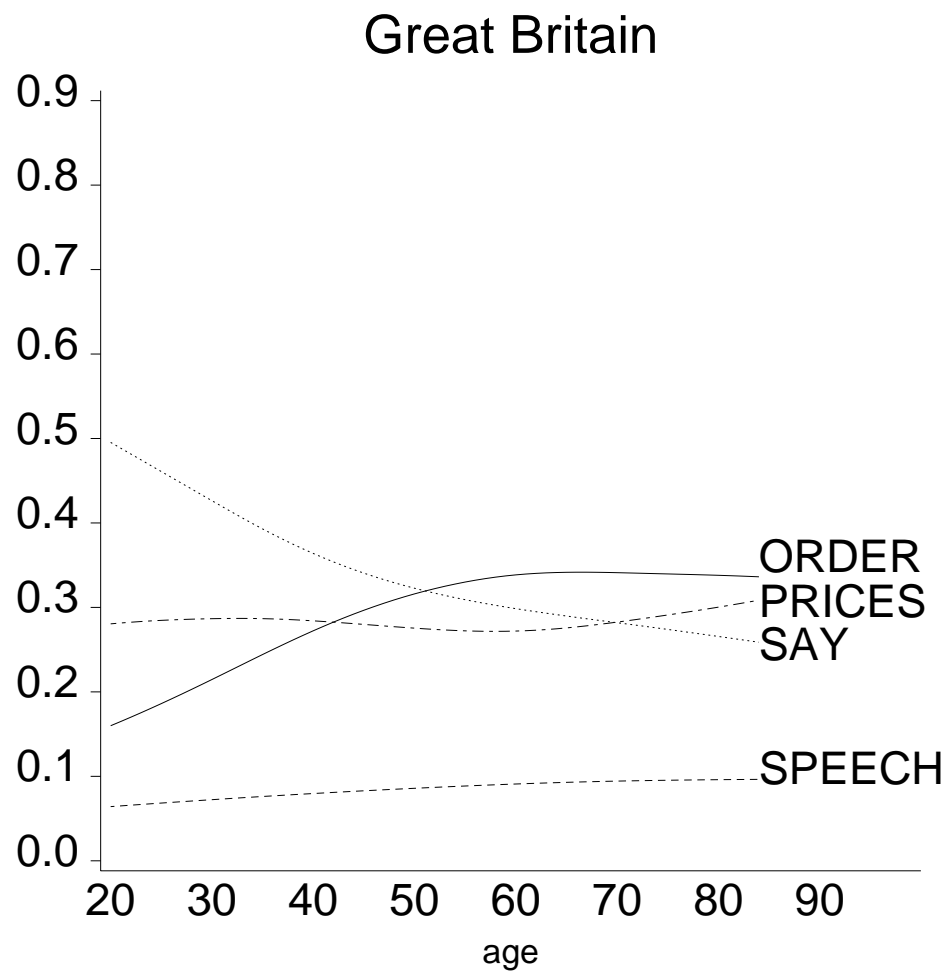
Francis, Dittrich, Hatzinger, Penn(2003) used a paired comparison approach with non-linear effects of age for cross sectional data.

$$\ln p_{i,(jk)} = \alpha_{i,(jk)}^* + 2\lambda_j + 2 \sum_{q=1}^M f_{jq}(x_{q,i}) + \sum_{q=M+1}^Q \beta_{jq} x_{q,i}$$

$p_{i,(jk)}$ is the probability of ranking item j above k for respondent i.

Such models can be extended to deal with multiple time points, but temporal dependence is complex.

... Value changes in society - cross-sectional results



... Value changes in society - Inglehart scales

- ✿ Many data sets
- ✿ World values survey 1999-2001 is sweep 4 - three earlier surveys.
Previous sweeps 1981-84, 1990-1993, 195-1997
- ✿ International social science program ISSP only certain years.
- ✿ Eurobarometer
- ✿ British Household Panel Survey - provides longitudinal data on Inglehart.
First five sweeps of study.
- ✿ European values survey 1990, 1981, 1999 www.europeanvalues.nl
- ✿ Northern Ireland life and times survey

We aim to use BHPS data and other panel data to collect longitudinal data on partial ranks, enabling us to explore generational and age effects.

... Value changes in society - Other directions

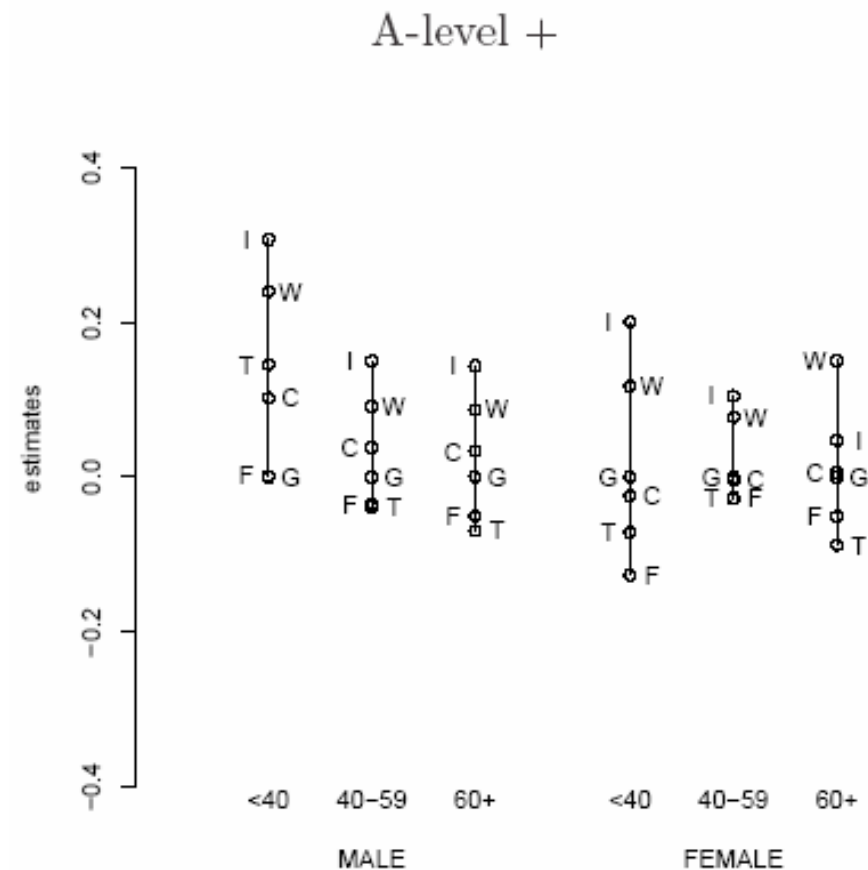
Other forms of value change are also worthy of interest:

1. Attitudes to environmental danger

(BHPS and other data sets)
analysis of multiple likert scales
over time.

Cross sectional analysis -
Dittrich et al(2005)

Note the difference between young GB
males and females in the relative
ranking of the danger of genetic
modification (G).



... Value changes in society - Other directions

Attitudes to gender

Six questions asked biennially: each in the form of a Likert scale:

- Q1: 'All in all, family life suffers when the woman has a full-time job.'
- Q2: 'A woman and her family would all be happier if she goes out to work.'
- Q3: 'Having a full-time job is the best way for a woman to be an independent person.'
- Q4: 'A husband's job is to earn money; a wife's job is to look after the home and family.'
- Q5: 'Children need a father to be as closely involved in their upbringing as the mother.'
- Q6: 'Employers should make special arrangements to help mothers combine jobs and childcare.'

... Patterns and pathways of criminal behaviour

... Patterns and pathways of criminal behaviour

- Early work in criminology in the 1970s attempted to classify a **criminal** - thus an offender might be judged a professional fringe violator, or a robber, or trickster (Gibbons 1972, Prentice Hall). An offender then has that label throughout their career. However, very little of this work was based on real life data, and difficult to classify new offenders to a class.
- We adopt a **developmental** approach - can we identify **types of criminal activity** in **distinct age regions** of an individual's history?
- Allows the development of an offender from one crime type to another. Criminologically, follows philosophy of Sampson and Laub (1993, Harvard UP) of **pathways through crime**.

... Patterns and pathways of criminal behaviour

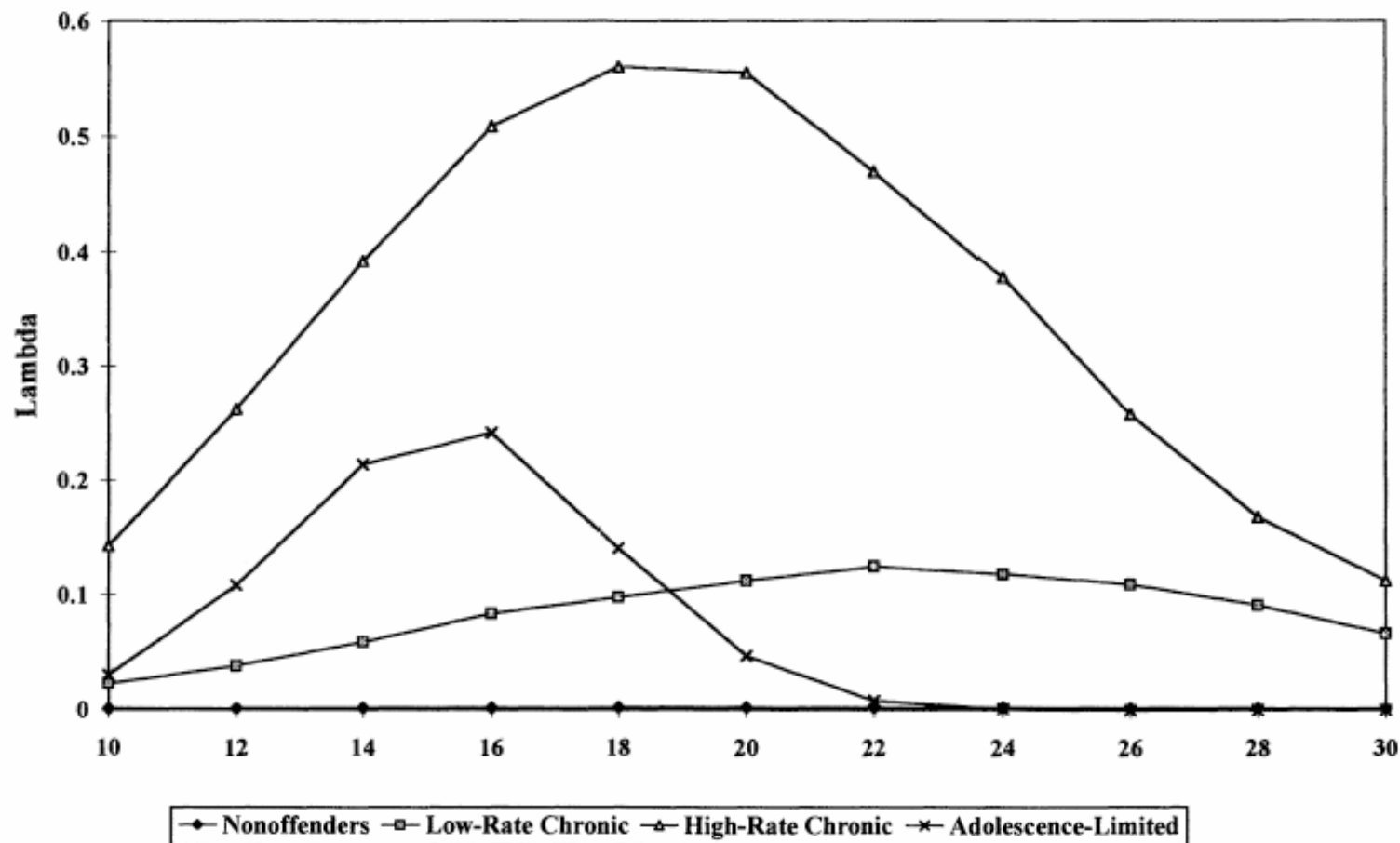


FIG. 1.—Predicted conviction rates by age for the London four-class model

Most work in US has focussed on the amount of offending over time.

Poisson latent class models have been used to model frequency of offending, for small samples of offenders.
(eg d'Unger et al, 1998, Am J Soc)

... Patterns and pathways of criminal behaviour

We are concerned with both **quality** and **quantity**.

A simplified criminal history of a typical male offender is shown below:

Age	14	17	20	22
Offences	Bicycle stealing	Shoplifting; Carrying offensive weapon	Fraud; Petty theft	Fraud; Petty theft; Receiving stolen property

We would like, for example, to determine whether **bicycle stealing** and **shoplifting** tend to co-occur in this cohort, whether **fraud** and **receiving stolen property** co-occur, and at what ages these offences are most prevalent.

We plan to use cohort data from the England and Wales Offenders Index.

Offence categories need to be simplified, reducing the offence codes to 73 major offences, after combining categories and eliminating offences with less than ten occurrences in the whole cohort (Francis et al, 2004 EuroJCrIm).

... Patterns and pathways of criminal behaviour

We are concerned with offending patterns in fairly short time windows. We will be defining a time window or **offence strip** of size h years and examining offending within that window.

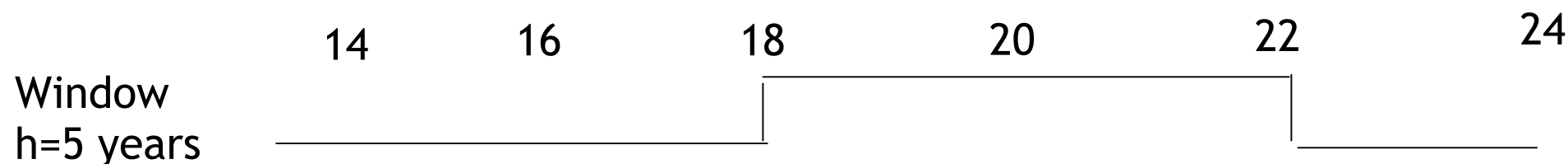
We are concerned with the **nature and variety** of offending, so within the window we define an prevalence matrix O_{ij}

$O_{ij} = 1$ if offender i is convicted for offence j within the offence strip or window
 $O_{ij} = 0$ otherwise.

We exclude strips with no offences and with all $O_{ij} = 0$

... Patterns and pathways of criminal behaviour

So, for $h=5$, we can for example look at offending between ages 18 and 22 inclusive, centred on age 20.



age	14	17	20	22
Offences	Bicycle stealing	Shoplifting; Carrying offensive weapon	Fraud; Petty theft	Fraud; Petty theft; Receiving stolen property

... Patterns and pathways of criminal behaviour

Various approaches have been considered.

- a) A simple latent class analysis with non-overlapping strips. This is equivalent to dividing up the time axis into disjoint age strips, and then clustering the binary measures using a likelihood based approach. Francis et al (2004) Euro J Crim.
- b) A development of a) with overlapping strips of fixed width, centred on actual ages of conviction. Local likelihood methods can be combined with latent class analysis.
- c) A multivariate latent markov model based on a), modelling transitions between latent classes between age strips as well as latent class membership. (Bartolucci, Francis, Pennoni, 2005)

... Patterns and pathways of criminal behaviour

We are interested in

- a) The **number of types** of activity. Does the repertoire of possible offending types broaden as offenders age?
- b) the **nature of the types of activity** - do these change for different cohorts, over time, and over age?
- c) The **transitions** from one crime activity type to another. Do offenders move from car crime to drug-related crime? Or from non-violent property to violent property acquisition?
- d) **When** the transitions occur. While we understand that every offender will make their own transitions, there will be particular ages at which a transition is most likely to occur.

... Patterns and pathways of criminal behaviour

We use the Offenders Index cohort data to investigate these ideas.

A complete criminal history of an approximate one in thirteen sample of all offenders born in 1953, 1958, etc up to 1978.

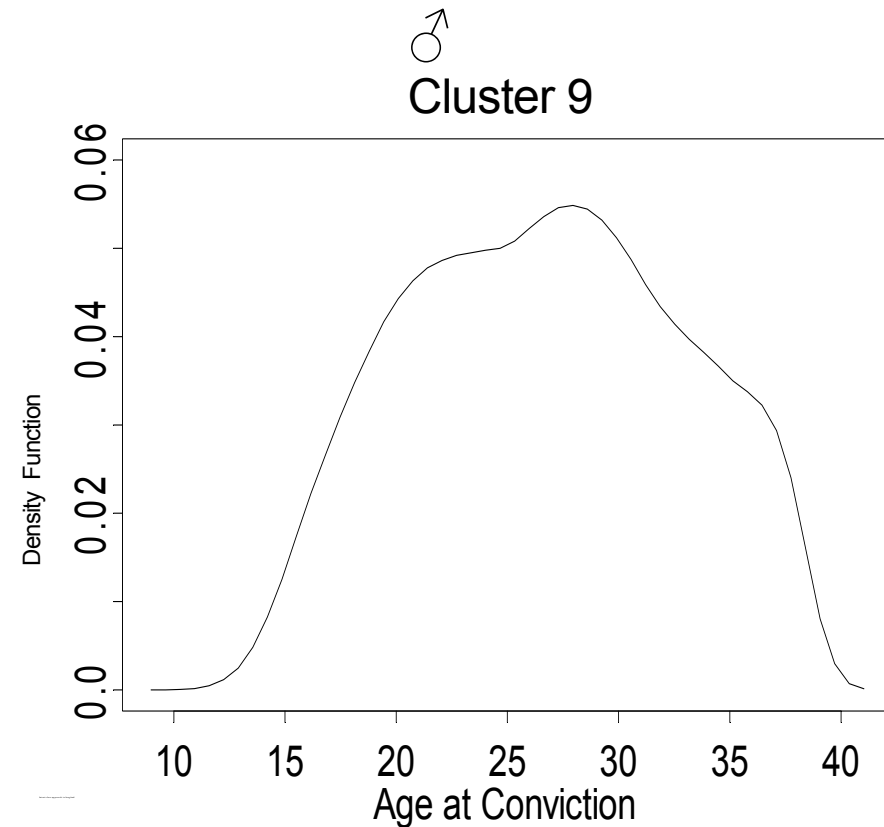
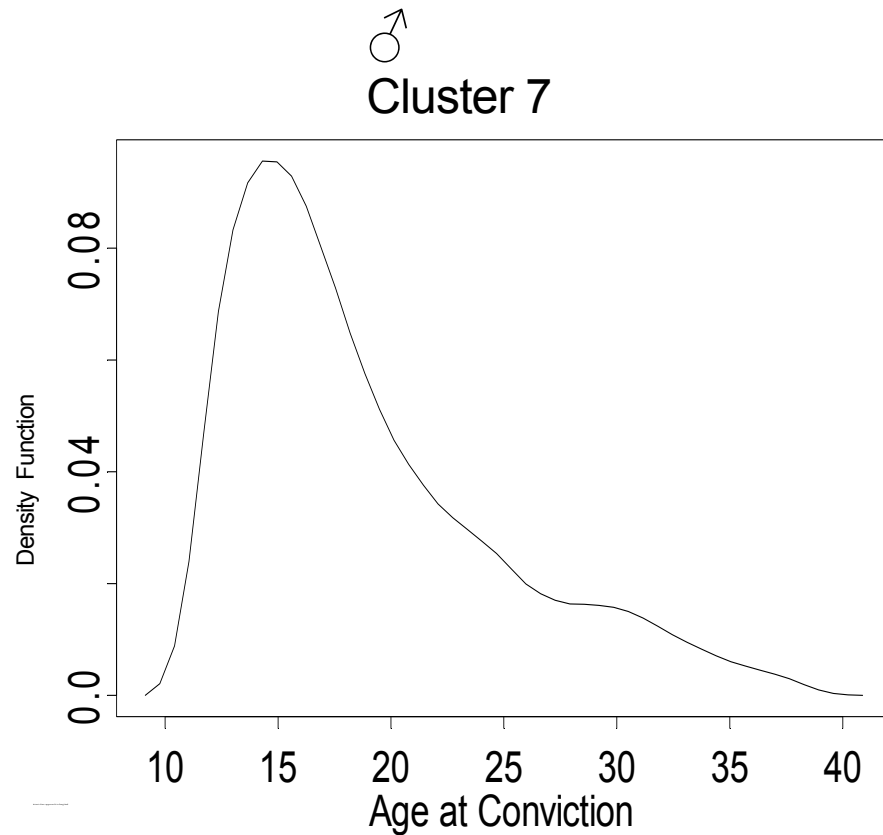
This allows varying amounts of follow-up time for each cohort.

HYPOTHETICAL EXAMPLE of dynamic clusters for each age-group:

MALES	< 16	16-20	21-25	26-30	31-35	36-40	41-45	46-50
1953	3	10	9	8	5	5	4	3
1958	4	12	9	9	6	5	4	X
1963	Etc.						X	X
1968						X	X	X
1973					X	X	X	X
1978				X	X	X	X	X
1983	Possible	Possible	X	X	X	X	X	X
1988	Possible	X	X	X	X	X	X	X

... Patterns and pathways of criminal behaviour

Early results Males 1953 cohort:



males
1953
birth
cohort

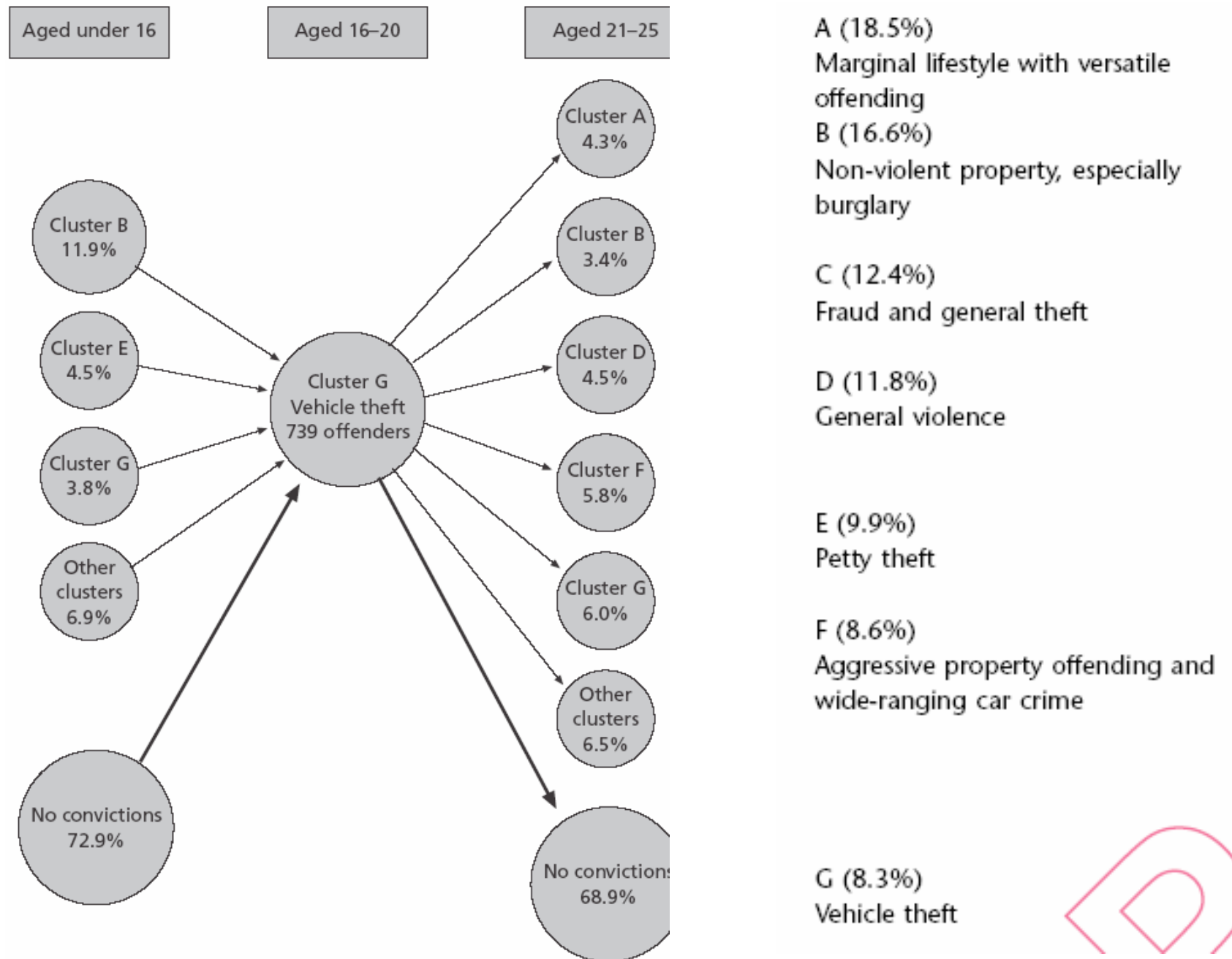


Figure 4 Offending pathways of male cluster G (vehicle theft), aged 16-20 years

... longitudinal modelling of child development data

... longitudinal modelling of child development data

Final substantive project – on child development.

Substantive aim

Develop methods for testing relative changes in cognitive skills over time.

Modelling strategy

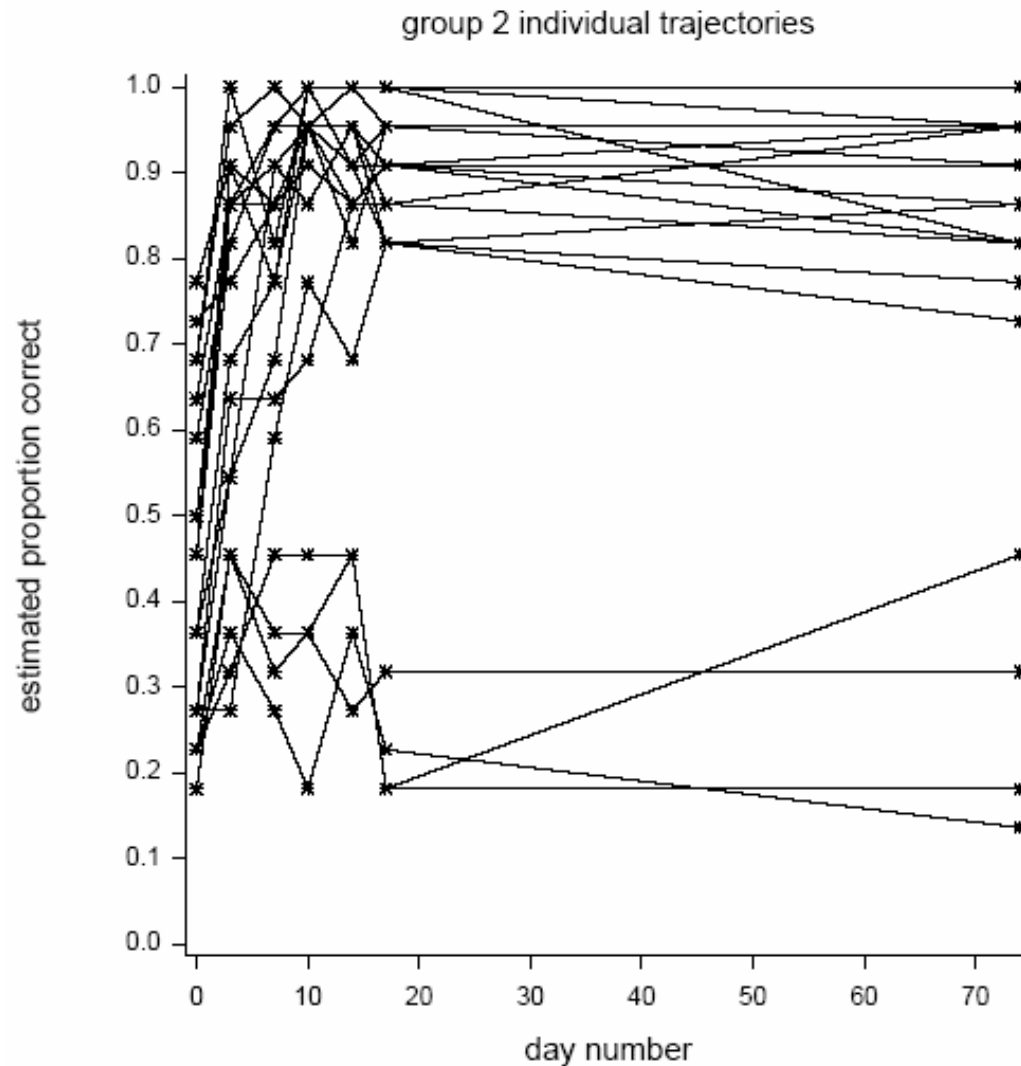
Development is a continuous-time process
But we need to fit models to discrete-time data.

... longitudinal modelling of child development data

What are we actually going to do?

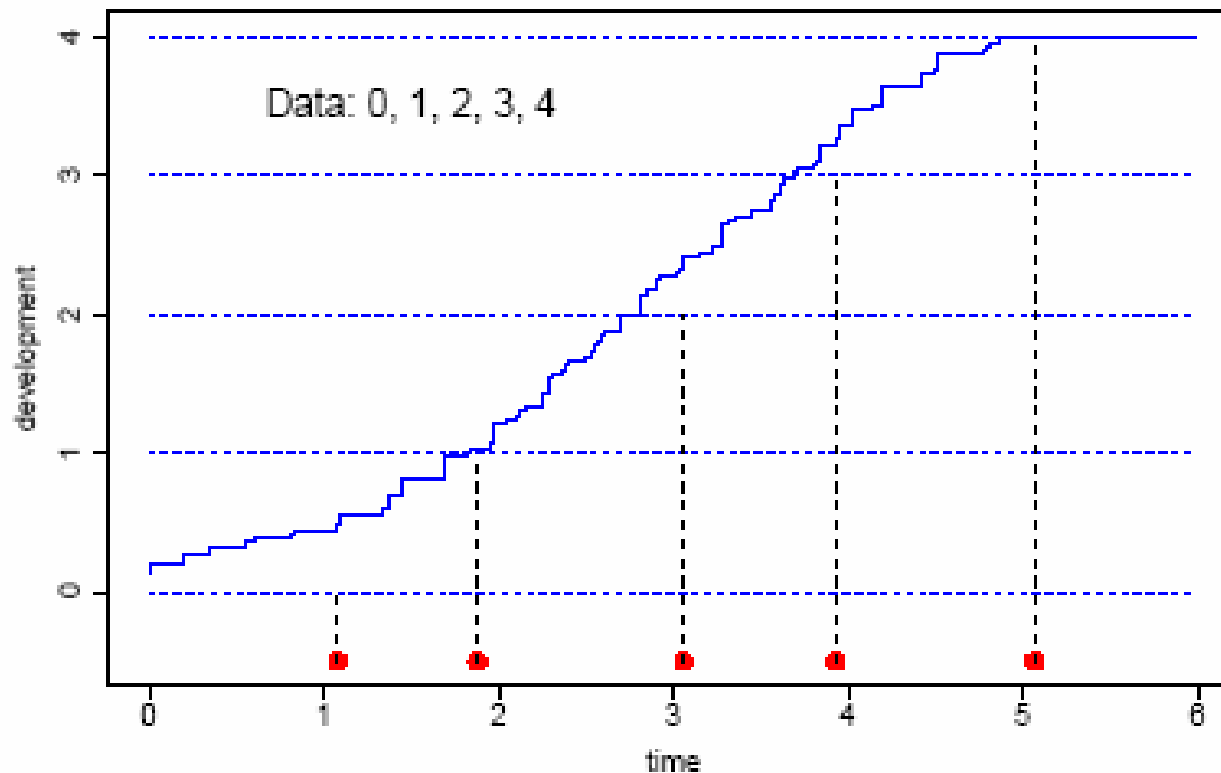
1. recruit $n = 50$ children at time zero (stage of development, not calendar age)
2. follow-up at times $t = 1; \dots, m \approx 5$. Microgenetic studies.
(in practice follow-up times vary between children)
3. administer $r \approx 20$ binary-outcome tests to each child at each follow-up time
4. how to combine binary test outcomes into an index of development?

... longitudinal modelling of child development data



... longitudinal modelling of child development data

1. Latent longitudinal development profile, $S(t)$, realised independently for each child.



... longitudinal modelling of child development data

Extensions: Some studies use two different batteries of tests:

- (a) postulate two latent development processes $S(t)$ and $S^*(t)$;
- (b) $S(t)$ and $S^*(t)$ correspond to qualitatively different aspects of development;
- (c) interested in understanding dependence (and, ideally, causal pathway) between $S(t)$ and $S^*(t)$.

Some methodological themes

- ▶ longitudinal data
- ▶ non-linearity
 - ▶ e.g., generalized linear
 - ▶ e.g., additive/multiplicative effects
- etc.
- ▶ complex random (or latent) effects
 - ▶ e.g., *crossed* effects $u_i + v_j$
 - ▶ e.g., effects in *competition*, $u_i - u_j$
 - ▶ *multiplicative* effects, e.g.,
$$g[E(y_{ij})] = (x_{ij}^T \beta + u_i) \exp(v_i)$$
- etc.

Problem: likelihood for nonlinear models with complex random effects often entails high-dimensional integrals

Standard approaches to approximating the likelihood:

- ▶ Monte Carlo methods, e.g., MCMC, simulated likelihood
 - ▶ dependence on random number sequence
 - ▶ demands expert tuning/diagnosis
- ▶ Laplace formula for integrals
 - ▶ persistent bias
 - ▶ biased standard errors

- └ Handling complex random effects
- └ Composite likelihood methods

A different approach: instead of the full likelihood, use a *pseudo-* or *composite* likelihood

- ▶ based on selected low-dimensional marginal views of the data
- ▶ replaces high-dimensional integral with many low-dimensional ones
- ▶ valid large-sample inferences
- ▶ some loss of statistical efficiency
- ▶ requires problem-specific ‘tuning’

e.g., Cox and Reid (2004, *Biometrika*)

- └ Handling complex random effects

- └ Composite likelihood methods

Aims:

- ▶ general guidance on
 - ▶ loss of efficiency
 - ▶ tuning

in some of the most commonly-occurring social science contexts (crossed-effect GLMs, etc.)

- ▶ empirical comparison with ‘standard’ approaches
- ▶ implementation as a fully documented *R* package, at least for all GLMMs.